PCT
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06T 11/00 | A1

(11) International Publication Number: WO 97/03416

(43) International Publication Date: 30 January 1997 (30.01.97)

(21) International Application Number: PCT/US96/11221

(22) International Filing Date: 8 July 1996 (08.07.96)

(30) Priority Data:
08/499,934 10 July 1995 (10.07.95) US

(81) Designated States: JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(71) Applicant: DAVID SARNOFF RESEARCH CENTER, INC. [US/US]; 201 Washington Road, CN5300, Princeton, NJ 08543-5300 (US).

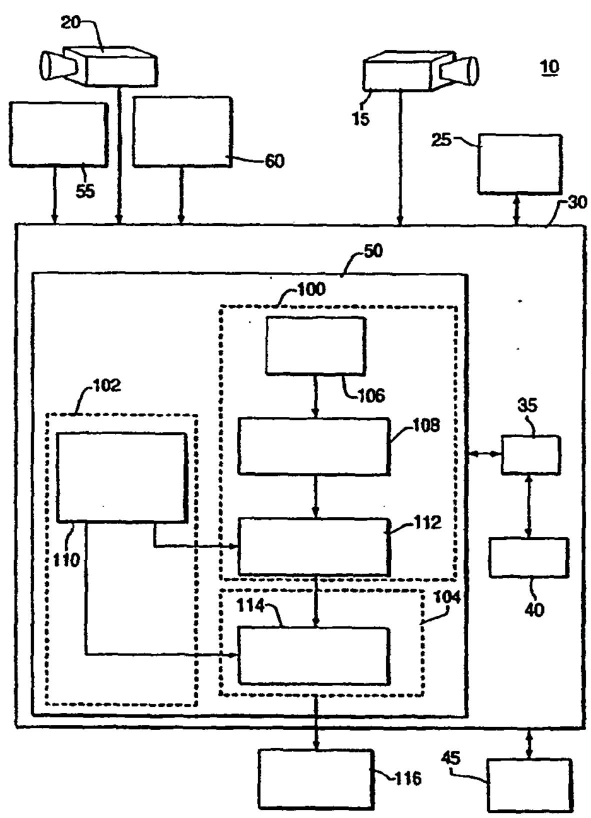
(72) Inventors: KUMAR, Rakesh; 64 Woodland Way, Dayton, NJ 08810 (US). ANANDAN, Padmanabhan; B-2 Carver Place, Lawrenceville, NJ 08648 (US). BERGEN, James, R.; 57 E. Prospect Street, Hopewell, NJ 08525 (US). HANNA, Keith, James; 212 John Street, Princeton, NJ 08542 (US). WILLIAMS, Kevin; 1213 Madison Drive, Yardley, PA 19067 (US). TINKER, Mike; 35 Highland Drive, Yardley, PA 19067 (US).

(74) Agent: BURKE, William, J.; David Sarnoff Research Center, Inc., 201 Washington Road, CN5300, Princeton, NJ 08543-5300 (US).

(54) Title: METHOD AND SYSTEM FOR RENDERING AND COMBINING IMAGES

(57) Abstract

An image processing system (10) for imaging a scene (15) to mosaic, selecting a new viewpoint of the scene, and rendering a synthetic image from the mosaic of the scene from that new viewpoint (112). The synthesized image is then combined with a second image (114). The combination of the second image and the synthetic image generates a composite image containing a realistic combination of objects in the second image and the scene. Using the system, a production set or other scene need only be created once, then imaged by the system. Thereafter, through image processing, any view of the scene can be synthesized and combined with separately imaged performers, or other objects to generate the composite image. As such, a production set or other scene can be repetitively reused without recreating the physical scene.



METHOD AND SYSTEM FOR RENDERING AND COMBINING IMAGES

The invention relates to image processing systems and, more particularly, to an image processing system and a concomitant method that derives a synthetic image of a scene from a mosaic of images and combines the synthetic image with a separately generated second image to form a realistic composite image having objects of the second image appear in the synthetic image.

To produce a film or television program, the entertainment industry spends as much as fifty percent of a production budget on the creation of "sets" including repetitive set-up and tear-down of the sets. For the production of a typical television program, an inordinate amount of time and effort is expended to repetitively set-up and tear-down the sets. The extensive amount of time required to create and use production sets limits the availability of such sets and, consequently, limits the creativity and flexibility of script writers and producers. Furthermore, utilization of complex and/or large sets further increases the production cost of the program.

To decrease the costs associated with set utilization, attempts have been made at synthetically generating objects and scenes using computer graphics. However, these graphical techniques generally produce images that lack detail because, as detail is added to a computer generated image, the processing time and cost escalates dramatically. As such, computer generated graphics are presently relegated to crude depiction of three-dimensional objects and scenes. Furthermore, the lack of image detail causes the images to have unrealistic or synthetic appearance.

Heretofore, there has not been an image processing system capable of recording a production set and rendering various viewpoints of the set that enable the set to be physically created once and then electronically reused. Moreover, there has been not been a system capable of recording a scene, rendering any view of the scene as a synthetic image, then combining the synthetic image with a second image such that a composite image is formed that realistically contains objects of the second image in the synthetic image.

Therefore, a need exists in the art for a system that eliminates the need for repetitively creating, setting up, and tearing down production sets by electronically imaging and storing the production set for subsequent, repetitive use. Furthermore, a need exists for a system that images a scene such as a production set, electronically stores the scene, permits the imaged scene to subsequently be realistically viewed from any viewpoint, including a moving

FIG. 6 depicts a block diagram of an image composition process.

To facilitate understanding, identical reference numerals have been used, where possible, to designate identical elements that are common to the figures.

The invention is an image processing system and a concomitant method for recording a three-dimensional scene such as a production set and subsequently reproducing views of that scene from any viewpoint. The system then combines the reproduced view (a synthetic image) with a second image of, for example, "real" performers that are imaged separately from the scene to form a composite image. The system combines the second image with the synthetic image of the scene from a viewpoint of the camera used to create the second image. As such, the "synthetic" view of the scene tracks the motion of the camera used to create the second image. Thus, the composite image realistically depicts objects in the second image within the scene environment, e.g., performers acting within a production set. Generally, the image of the scene and the second image are video images, i.e. sequences of image frames each containing an array of pixels. As such, the composite image is a sequence of images forming video. In its broadest sense, this invention can be used to process and combine any digital representation of images including individual images such as still photographs, or moving pictures such as film images and video, or animation art, or any combination thereof.

FIG. 1 depicts a high level block diagram of the an image processing system 10 in accordance with the invention. Specifically, FIG. 1 depicts a combination of hardware implementing the system and process steps executed by the system hardware to attain system functionality.

The system 10 comprises a computer system 30, a scene image camera 15, a second image camera 20, at least one camera parameter sensor such as a camera orientation sensor 55 and/or a camera position location system 60, a video storage device 25, and one or more input and output devices for the computer system. The computer system may be a general purpose computer system having a central processing unit (CPU) 35 that is programmed by executing one or more programs residing in memory 50 (e.g., random access memory, floppy disk, hard disk and the like). The CPU functions in conjunction with well-known support circuits 40 such as read only memory, power supplies, co-processors, and the like. The computer system is controlled by conventional input devices and displays information via conventional output devices. The video storage device is an optional system component that is used depending upon whether the CPU can process video

to determine a viewpoint from which the scene should be viewed to produce the synthetic image. The system combines, at step 112, the mosaic representation with the camera parameters to derive (or synthesize) a synthetic image of the scene viewed from the camera position used to record the second image. The system then combines, at step 114, this synthesized image with the second image produced in step 110. The final image is displayed, at step 116, where this image (typically, a video sequence of images) contains the object of the second image within the scene. Importantly, as the second image camera view changes, e.g., camera pan and tilt, the synthetic image of the scene varies to match the camera motion. Consequently, the object of the second image appears as if it were imaged contemporaneously with the three-dimensional scene. As such, if the scene is a production set and the second image contains performers, the composite image contains the performers acting upon the production set.

Each of the processes that comprise the image processing system is individually discussed in detail below.

FIG. 2 depicts a hardware arrangement of camera(s) within a three-dimensional scene, illustratively studio or production set 200, that is used by the image rendering process to generate a mosaic representation of the scene. The mosaic is generated by a mosaic generation system (not shown) such as the two-dimensional mosaic system disclosed in commonly assigned U.S. patent application serial number 08/339,491 entitled "Mosaic Based Image Processing System" filed November 14, 1994 and incorporated herein by reference, or the three-dimensional mosaic system disclosed in commonly assigned U.S. patent application serial number ________, attorney docket number 11789, entitled "Method And System For Image Combination Using A Parallax-Based Approach" filed June 22, 1995 and incorporated herein by reference. Each of these mosaic generation systems would sufficiently function to render a useful mosaic. However, for best results, the three-dimensional mosaic generation technique is preferred.

As described in the docket 11789 application, given an existing 3D mosaic representing a three-dimensional scene and the pose (rotation, translation and zoom) of a new viewpoint with respect to that mosaic, the system can derive a synthetic image of the scene. As such, by capturing a scene using different cameras having different viewpoints of the scene, the system can synthesize images that are a view of the scene from viewpoints other than those f the cameras. Of course, a single camera can be used to image the scene from different locations and th mosaic generated from those images.

discloses an illustrative technique for producing mosaics using hierarchical direct registration. Furthermore, once the mosaic is generated, additional imagery can be added to the mosaic using conventional image merging and fusing techniques.

More specifically, as shown in FIG. 3, the various images recorded at each camera location are combined into a plurality of mosaics, i.e., one mosaic 300 for each grid point. To produce each of the mosaics, the camera is panned, tilted, rolled and zoomed at each grid point. As such, for each grid point, the system generates a mosaic containing image information with respect to four camera parameters. Since each mosaic is related to a specific three-dimensional location, the mosaics form an addressable three-dimensional array. Each mosaic is related to its neighboring mosaics by a parametric transformation 302 and a motion flow field 304. Since the production set being imaged is typically static, the motion flow field represents parallax motion of objects within the image, i.e., the three-dimensional geometry of the scene. This parallax motion is also referred to as the "shape" of the scene. In other words, as the camera moves from location to location, parallax causes objects in the scene to move relative to the background, e.g., as a camera pans, a foreground chair seems to move with reference to a background wall. By storing the mosaics as well as a parametric transformation relating one mosaic to the next, e.g., the motion of the background as the camera physically moves, and the motion flow field information, e.g., the parallax information representing the three-dimensional geometry of the scene, the entire scene can be recreated from any viewpoint and none of the three-dimensional geometry is lost. Thus, a given three-dimensional mosaic comprises an image mosaic representing a panoramic view of the scene and a shape mosaic representing the three-dimensional geometry of the scene.

The process used to generate a synthetic image view of the scene is known as "image tweening". This process warps each of the individual mosaics (e.g., mosaics 206, 208, and 210) to the location of the synthetic viewpoint (e.g., location 220). Thus, as each 3D mosaic is generated for each grid point, the 3D mosaic is stored in memory (mosaic storage 222) with respective to its associated grid point. Given a new viewpoint location, the mosaics are recalled from memory to generate a synthetic image representing the sc ne from the new viewpoint. Depending upon the complexity of the scene being imaged, the system may r call each of the 3D mosaics in memory or some subset of those mosaics, e.g., only recall those mosaics that are nearest the new view location. Using image tweening process 224, each

indicium of viewpoint is, for this illustrative example, at least one camera parameter and, typically, is a set of camera parameters that are corrected by the computer system. Using a set of corrected camera parameters, the second image is subsequently combined with the synthetic image to produce the composite image.

Typically, to facilitate the use of chroma-key techniques during the image compositing process, the object of the second image is positioned before a blue (or sometimes green) screen. In the illustrated room 402, the walls 404 are painted blue. To aid in camera position tracking, the walls contain a series of indicia 408 positioned about the top of the room 402, e.g., blue X's on a blue wall that lie outside the scene being imaged. Of course, the indicia may be located on the floor or any other location in the studio. These indicia may be identical to one another or unique. Furthermore, by using blue indicia on the blue walls, the indicia are easily extracted from the second image leaving only the object of the second image for combining with the synthetic image.

Preferably, these indicia 408 are continuously imaged by one or more "tracking" cameras 406 (one of which is shown) located atop the second image camera 20. Alternatively, as discussed in detail below, the indicia may be directly imaged by the second image camera as it produces the second image. In either case, once processed and combined with the synthetic image, the indicia form no part of the second image because the indicia, if they appear at all, are designed to be extracted from the second image by the image compositing process.

In addition to the tracking camera, the second image camera is outfitted with a zoom sensor 410 and a variety of camera orientation sensors 412. A high speed computer system 30 stores the output video from the tracking camera as well as the camera parameters. The tracking camera video signal may alternatively be stored separately in a video signal recording medium such a video tape. Likewise, the video signal (hereinafter referred to as the performance image or second image) from the second image camera is sent directly to the compositing process for real-time combination with the synthetic image or, alternatively, the video signal may be recorded onto a recording medium such as a video tape.

Camera position determination is accomplished using two steps; namely, in step 1, the system directly measures camera parameters to crud ly estimate the cam ra position, and, in step 2, the system uses the indicia images by the tracking camera (hereinafter referred to as the reference image) to refine (correct) th camera parameters of step 1. As the camera moves while producing the s cond image, the computer 30 records, in

representation of the studio, while the second method 512 uses an image representation of the studio.

For the symbolic representation of the studio, the position of each landmark is premeasured, at step 514, in absolute terms with reference to a fixed coordinate system, e.g., having the origin of the coordinate system at the corner of the studio. Using the tracking camera image(s) provided in step 524, the method determines the pose of the camera, i.e., the rotation, translation and zoom parameters of the camera relative to the reference coordinate system. The camera pose is computed at step 516 by first estimating the pose using the measured camera parameters, then refining (correcting) the estimate using a pose estimation process. Camera pose determination and estimation using indicia of scene orientation is a well-known technique. See, for example, Kumar et al., "Robust Methods for Estimating Pose and a Sensitivity Analysis", CVGIP: Image Understanding, Vol. 60, No. 3, November, pp. 313-342 (1994). Using this technique and given the correspondence between indicia in an image, the technique determines rotation and translation matrices that map a reference or "world" coordinate system to a "camera coordinate system. The technique described in this paper is applied to the indicia on the studio walls such that the camera pose first estimated by the sensors is refined to compute accurate camera parameters. The system iterates the camera pose through the levels of the image pyramid representation of the reference image until a sufficient degree of accuracy is achieved. At step 520, the system outputs the corrected camera pose.

For the image representation of the studio, at step 502, the entire studio is imaged from various, known locations within the room (e.g., using a grid pattern) and the video signal from the tracking camera is stored as a series reference images. At step 504, these reference images of the indicia are stored with reference to an array of camera positions as measured by the camera position location system and orientation sensors. The array maps the camera positions, as measured at each grid location with the grid pattern, to a specific view of the indicia in each reference image. In this manner, given, at step 506, a camera pose (e.g., a set of camera parameters representing the camera's rotation, translation and zoom with respect to a known reference coordinate system), the system recalls, at step 508, a particular view of the indicia, i.e., the system recalls a particular reference image. Typically, to p rmit rapid computations using the images, each reference image is stored as an image pyramid. Image pyramids are well-known in the art for their use in representing a single image as a series of levels where each level has a lesser resolution than a previous level. Generally, image pyramids are formed by

As such, the indicia image is, in essence, embedded into the second image and the reference images are prerecorded using the second image camera. In this alternative approach, the field of view of the second image camera is broad enough to image the indicia as well as the object of the second image. The process used to compute the offset parameters remains the same as described above, e.g., the reference image is warped into alignment with the landmark image (now a portion of the second image) and computing the offset parameters based on the degree of image warping used to align the reference and indicia images. Subsequently, during image composition, the indicia can be masked from the final image. Various techniques for camera orientation indicia tracking and utilization are discussed in commonly assigned U.S. patent applications serial number 08/222,207, filed March 31, 1994; serial number 08/380,484, filed January 30, 1995; and serial number 08/456,020, filed May 31, 1995. The disclosures of each of these patent applications are herein incorporated by reference.

Additionally, some of the parameters do not have to be as accurately determined as other parameters. As such, the computation of the absolute camera position can be simplified. The simplification involves smoothing some of the camera parameters such as zoom, pan, tilt, and roll, while other parameters such as position parameters (x,y,z) are corrected using one of the landmark imaging techniques described above. In this manner, the indicia imaging techniques need only be applied to three variables and the four other variables are mathematically computed using averaging or interpolation of the measured parameters.

The foregoing sections discussed generating a three-dimensional mosaic of a scene (e.g., a production set, background image and the like) and also generating a second image as well as indicium of viewpoint associated with the second image. Although the foregoing discussion has focused upon generating the second image using the tracking process, the second image can be generated by any number of sources including computer graphics, animation art, a second synthesized image rendered from a second 3D mosaic, historical films, photographs, and the like. This section describes the process by which the inventive system combines the second image with a synthesized image extracted from the three-dimensional mosaic of the scene. The result is a realistic composite image depicting the object of the sec nd image within the scene.

FIG. 6 depicts a block diagram of the image compositing process 104. As described above, the image rendering process 100 provides a synthetic image, on path 606, from a plurality of images 602 and 604 of a scene. The

Specifically, to accommodate various light levels when producing the composite image, three different light levels are used while generating the individual images for the mosaics. In effect, three mosaics are produced for each grid point, one mosaic for each light level. During image synthesis, the synthesized image is produced by interpolating pixel luminance of the three mosaics for each pixel in the synthesized image. The weighting used in the interpolation is user defined, such that a system operator can adjust the degree of inclusion or exclusion of one of the three mosaics to achieve a lighting level that matches the illumination in the performance image.

In addition to the foregoing illumination compensation process, the system may use a number of image filtering and enhancement techniques to adjust the luminance level of a specific portion(s) of the image. For example, objects in the synthesized image having surfaces that are perpendicular to the direction of the light used to illuminate the object of the second image can be enhanced to appear to also be illuminated. In this manner, a synthesized image of a scene seems as if it were recorded contemporaneously with the second image.

Specifically, when illumination is projected onto a scene, structure that lies perpendicular to the illumination typically has contrast. One approach to synthesize a new illumination direction is to decompose an image into structure at different orientations, modify the structure at each orientation, and recombine the image to produce an altered output image. For example, an image can be decomposed into several gradient orientation and scale bands using an oriented Laplacian pyramid. This pyramid has the property that summing all the bands unchanged results in the original image. To synthesize the effect of a different illumination angle upon the image, those orientation bands perpendicular to the desired synthetic illumination angle are amplified, and those orientation bands parallel to the desired illumination angle are attenuated. To illustrate, an oriented Laplacian pyramid can be computed from an input image resulting in a set of four filtered images at several scales. The cosine of the angle between the filter orientation that produced each filtered image and the desired synthetic illumination is then computed. The filtered image is then multiplied by this gain. This is accomplished for each orientation at each scale. The images are then recombined to produce an output image. An adjustment allows an operator to cycl through different illumination settings until the desired synthetic illumination is attained.

Additionally, the gains on the riented filter outputs can be adaptively adjusted depending on the image structure, rather than a single gain for each filtered image. This allows for more accurate illumination synthesis since

WE CLAIM:

1. A method for image processing comprising the steps of:

generating a mosaic containing a plurality of first images of a scene;

generating a second image;

producing an indicium of viewpoint while generating said second image;

rendering, in response to said indicium of viewpoint, a synthesized image from said mosaic; and

compositing said synthesized image with said second image to produce a composite image.

2. The method of claim 1 wherein said mosaic is a plurality of mosaics.

3. The method of claim 1 wherein said step of generating said second image includes the steps of rendering said second image from a second mosaic.

4. The method of claim 1 wherein said indicium of viewpoint is at least one camera parameter.

5. The method of claim 1 wherein said step of generating a second image further comprises the step of imaging a performance before a background having a color that can be removed from the second image using a chroma-key technique.

6. The method of claim 1 wherein said step of producing an indicium of viewpoint further comprises the step of recording at least one camera parameter selected from the following group of camera parameters: three-dimensional position, pan, tilt, roll and zoom parameters.

7. The method of claim 1 wherein said step of producing an indicium of viewpoint further comprising the steps of:

providing a plurality of indicia proximate an area where a performance is imaged as said second image; and

imaging said plurality of indicia as a reference image while generating said second image.

8. The method of claim 7 wherein said landmark imaging step further comprises the step of generating said second image using a camera.

9. The method of claim 7 wherein said landmark imaging step is accomplished using a tracking camera that is separate from a camera used to generate said second image.

10. The method of claim 7 further comprising a step of imaging a plurality of reference images representing said indicia viewed from a plurality of viewpoints.

11. The meth d of claim 10 further comprising a step of correcting said indicium of viewpoint using said plurality of reference images.

12. The m thod of claim 11 further comprising the steps of:

mosaic that represents a view of the scene corresponding to the indicium of viewpoint; and

compositing means, connected to said rendering means and said viewpoint indicium means, for combining said synthesized image and said second image to form a composite image.

20. The system of claim 19 wherein said mosaic generation means further comprises:

means for generating a plurality of mosaics, where each mosaic in said plurality of mosaics represents a panoramic view of the scene from a unique location.

21. The system of claim 19 wherein said viewpoint indicium means further comprises:

means for tracking a camera pose as said camera produces said second image; and

means, connected to said tracking means, for generating said camera pose as an indicium of viewpoint.

22. The system of claim 19 wherein said viewpoint indicium means further comprises:

means for producing said indicium of viewpoint in response to user commands.

23. The system of claim 19 wherein said compositing means further comprises:

means for combining said synthetic image and said second image using a chroma-key process.

24. The system of claim 21 wherein said camera pose includes indicium of viewpoint selected from the following group of camera parameters: three-dimensional position, pan, tilt, roll and zoom.

25. The system of claim 19 further comprising:

second mosaic generation means for generating a second mosaic of images representing a second scene; and

second rendering means, connected to said viewpoint indicium means and said second mosaic generation means, for producing said second image.

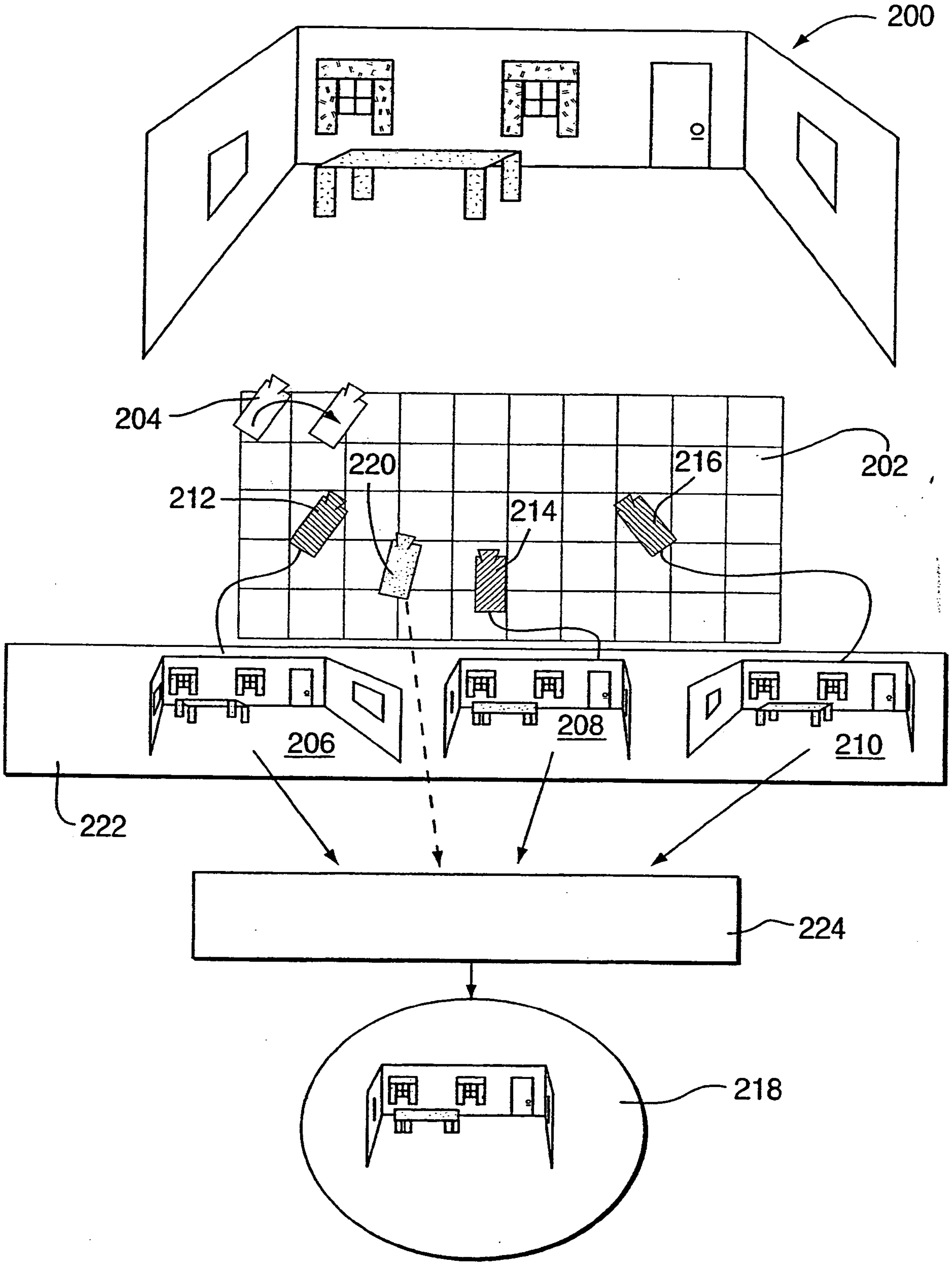


FIG. 2

BNSDOCID: <WO 9703416A1 I >

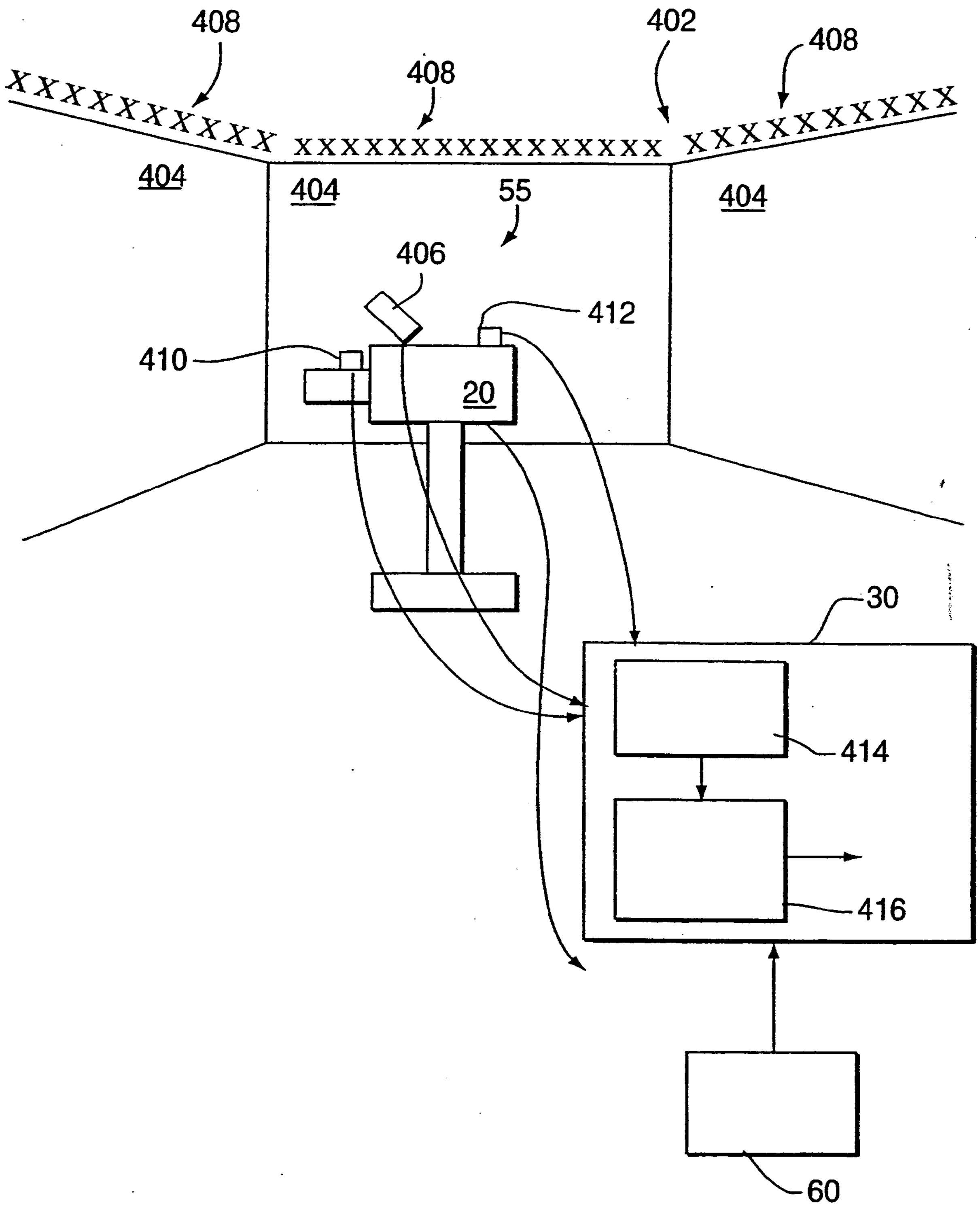


FIG. 4

BNSDOCID: <WO 9703416A1 I >

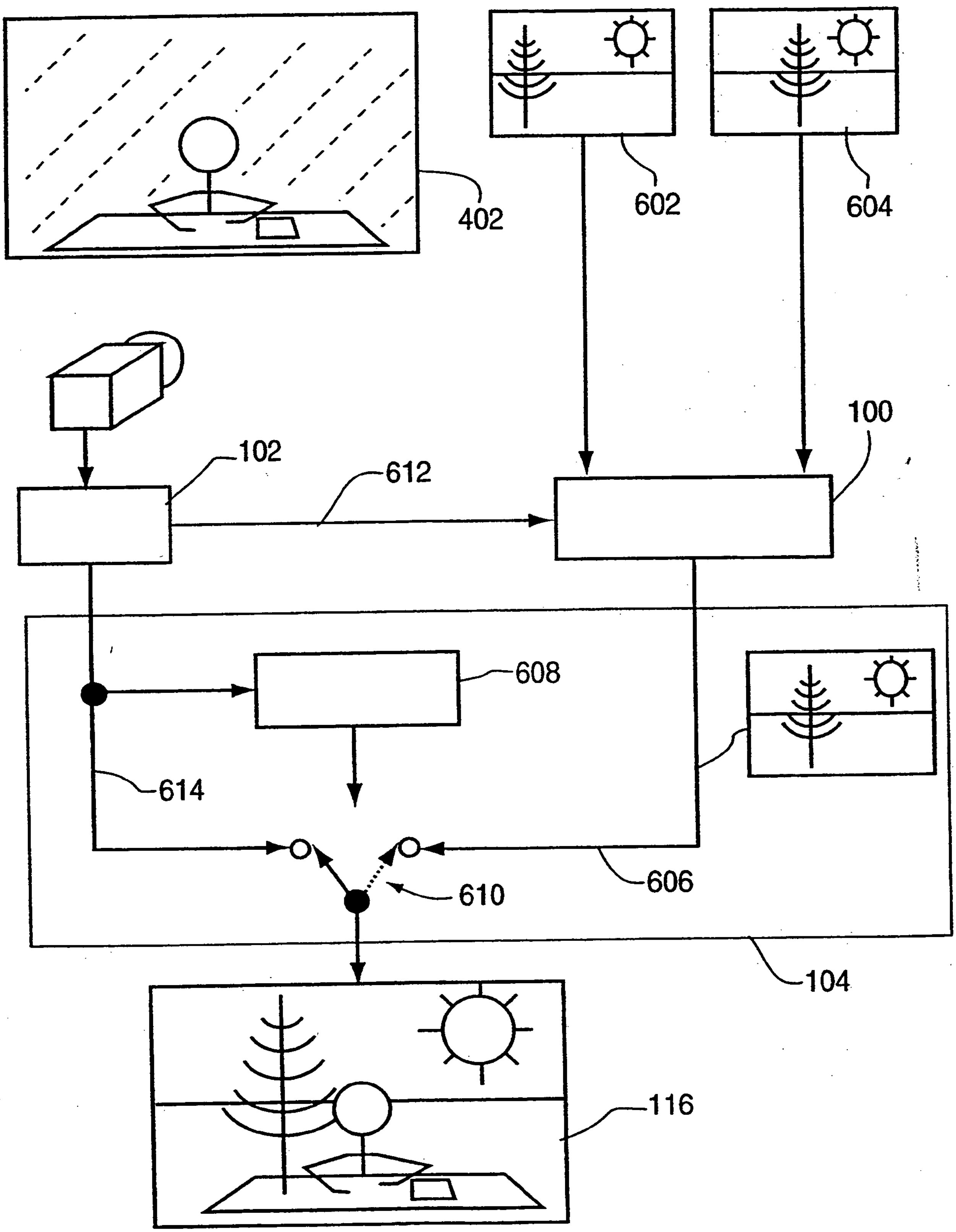


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/11221

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Proceedings of the 1994 IEEE Workshop on Motion of Non-Rigid and Articulated Objects, 11 November 1994, H. S. Sawhney, "Simplifying Multiple Motion and Structure Analysis Using Planar Parallax and Image Warping", pages 104-109, especially pages 104-107	12
Y	Proceedings Third International Conference on Computer Vision, 04 December 1990, R. Kumar et al., "Sensitivity of the Pose Refinement Problem to Accurate Estimation of Camera Parameters", pages 365-369, especially page 365	13, 21, 24
A	Proceedings of the IEEE Workshop on Visual Motion, 07 October 1991, K. J. Hanna, "Direct Multi-Resolution Estimation of Ego-Motion and Structure from Motion", pages 156-162	1-25
A	Proceedings Fourth International Conference on Computer Vision, 11 May 1993, R. Cipolla et al., "Robust Structure from Motion Using Motion Parallax", pages 374-382	1-25
A	US, A, 4,758,892 (BLOOMFIELD) 19 July 1988, abstract	1-25
A	US, A, 4,797,942 (BURT) 10 January 1989, abstract	1-25